Projet Régression

Prédire le prix d’une voiture d’occasion

**Equipe :**

Axel Arcidiaco

Briand Bakouzou

Etienne Wagner

Jean-Paul Sossah

# Sommaire

Contexte

Planification

Guide utilisateur

Documentation technique

Difficultés rencontrées

Perspectives d’évolutions

Conclusion

Bilan de groupe

Bilans personnels

# Contexte

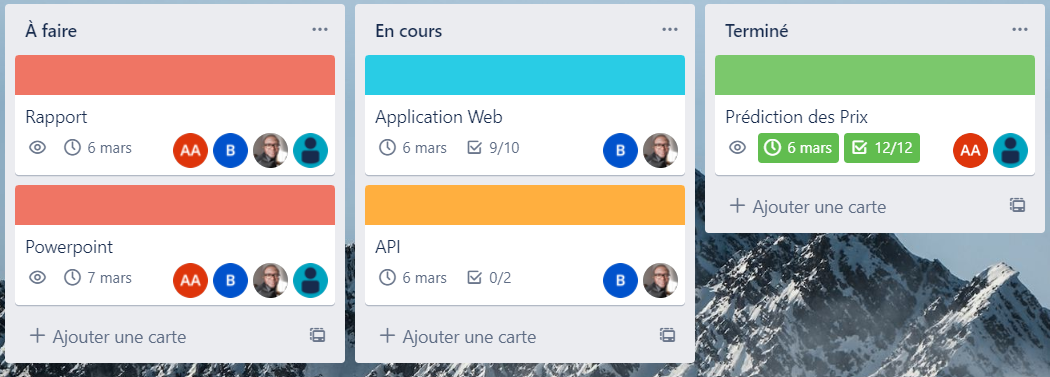
Comme le marché de l'occasion automobile est complexe et dépend de plusieurs paramètres tels que l'année d'immatriculation, la marque, la fiabilité et le carburant, ainsi que la cote établie en fonction de la popularité du véhicule, nous avons été chargés par notre client, leader du marché des véhicules d’occasion, de réaliser une application permettant de prédire le prix du marché automobile en fonction de différents critères donné par l’utilisateur.

Pour répondre aux demandes de notre client, nous avons donc créer une application web avec une expérience utilisateur optimale, c’est à dire simple d’utilisation et facile à comprendre, pour des concessionnaires automobiles très peu formés aux outils numériques ainsi qu’une API pour permettre à une entreprise tierce de développer une application mobile.

# Planification

Pour la réalisation de ce projet, nous avons utilisé l’outil de gestion de projet en ligne Trello avec lequel nous avons pu effectuer le découpage et la répartition des tâches du projet.

Ces tâches était la préparation des données auxquelles nous avions accès, l’implémentation et le test de différents modèles de régression avant de choisir lequel d’entre eux était le mieux adapté à nos besoins, la conception et le développement de l’application web, et finalement le développement de l’API.



Nous avons aussi utilisé les applications Microsoft Teams et Discord pour partager des fichiers et éléments de codes entre nous.

Guide Utilisateur

Documentation technique

Préparation des données

La première étape de notre projet a été la préparation des données qui nous avais été fournis afin de pouvoir les exploiter avec des modèles de machine learning, dans ce que des modèles de régression tel que la régression linéaire, pour prédire le prix de voitures de location.

Pour préparer les données qui nous avons été fourni, nous avons fait le choix de retirer la colonne “New\_Price” de nos datasets d’entrainement et de test car celle-ci contenait une large majorité de valeur nulles contre un faible nombre de valeurs non nulles (5195 valeurs nulles pour 824 valeurs non nulles dans le cas du dataset d’entrainement) et que nous avons jugé que remplir les valeurs manquantes de cette colonne en utilisant les valeurs existantes

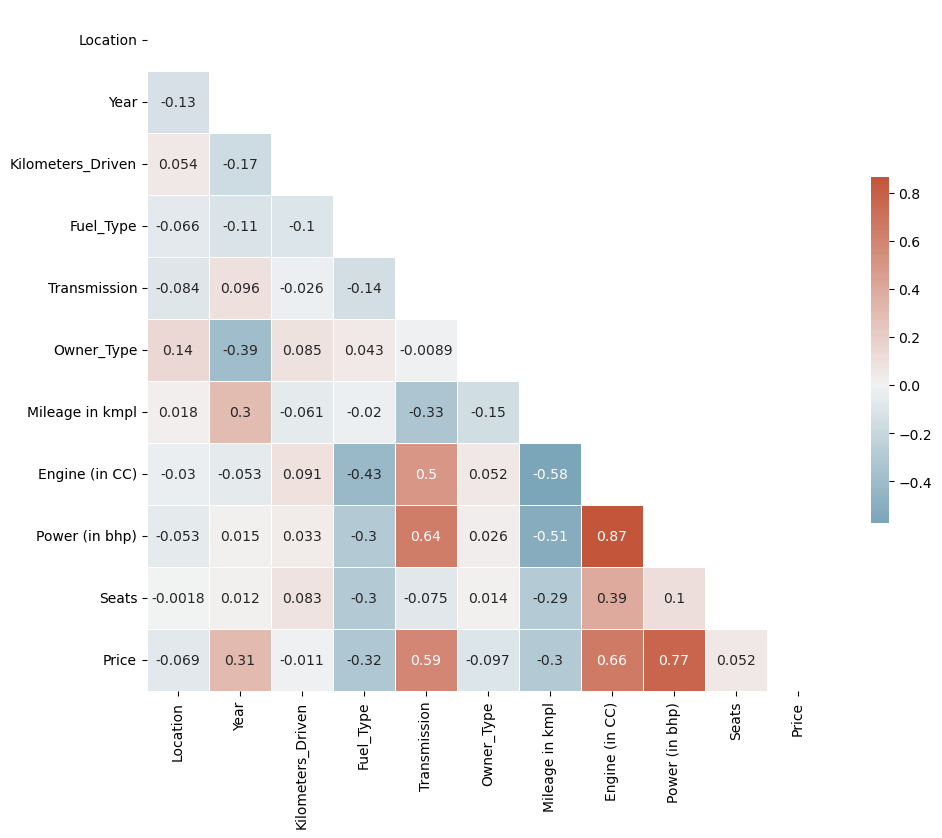
Après avoir retiré la colonne “New\_Price” de nos datasets, nous avons retiré des datasets les autres valeurs nulles qu’ils contenaient avant de transformer les données en convertissant les données de la colonne “Mileage” dans la même unité de mesure et en retirant des données des colonnes “Mileage”, “Engine”, et “Power” le symbole de leurs unités de mesure.

Une fois cette préparation de nos données terminée, nous avons alors pu passer à l’implémentation et à l’évaluation de divers modèles de régression.

Modèles de régression

Après avoir terminé la préparation de nos données en retirant de nos datasets les valeurs nulles qu’ils contenaient et en adaptant certaine de nos données pour qu’elles soient exploitable par nos modèles de régression, nous avons entamé une phase de test et d’évaluation de plusieurs modèles de régression pour déterminer lequel ou lesquels seraient le mieux adapté à notre problématique de prédiction des prix du marché de l’occasion automobile.

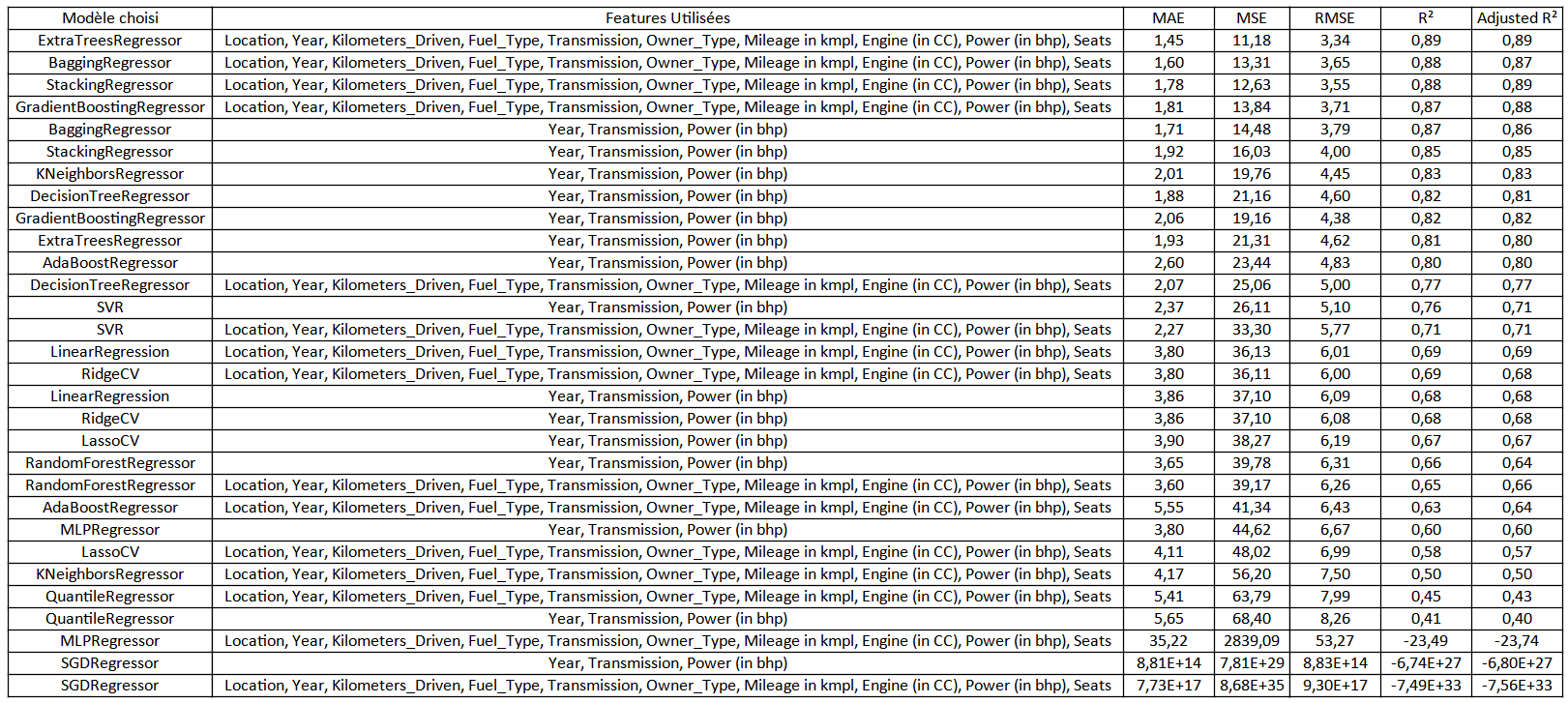
Pour ce faire la première étape que nous avons effectué a été de générer une matrice de corrélation pour déterminer qu’elles étaient les features les plus importante parmi nos données pour prédire le prix d’une voiture d’occasion.



Comme on peut le voir sur la matrice de corrélation ci-dessus, on peut constater que les features possédant un grand taux de corrélation avec le prix d’une voiture d’occasion sont : “Power” qui représente la puissance du moteur d’une voiture, “Engine” qui représente le volume du moteur d'une voiture, “Transmission” qui représente le type de transmission (manuel ou automatique) d'une voiture, et “Year” qui représente l’année de production d’une voiture.

De plus, on peut voir que les features “Engine” et “Power” possèdent un taux de corrélation important ce qui nous indique que nous ne sommes pas obligés de nous servir de ces deux caractéristiques en même temps.

Après avoir identifié les features les plus importantes pour la prédiction du prix d’une voiture, nous avons testé différents modèles de régression pour déterminer lequel serait le mieux adapté pour répondre à nos besoins.



Application Web

API

Difficultés rencontrées

Perspectives d’évolutions

Conclusion

Bilan de groupe

Bilans personnels